Supplementary data

I. Supplementary Methods

Image Segmentation and Radiomics Feature Extraction Methodology

Volumes of interest (VOIs) of the bladder tumor were semi-automatically segmented using the GrowCut segmentation method implemented in 3D Slicer software. GrowCut is an interactive region growing segmentation method.[1] Using the GrowCut method, VOI is initially delineated. Then radiologists can meticulously edit the boundary of the Regions of interest (ROIs) slice-by-slice by erasing or drawing the mask manually, improving the alignment of the ROIs with the tumor outlines.

In this study, a total of 718 radiomics features were extracted using the *PyRadiomics* platform implanted in the publicly available 3D Slicer software. The features could be divided into four categories: (a) first-order statistics features, (b) shape-based features, (c) statistics-based textural features, and (d) wavelet features. Detailed descriptions of the feature classes are presented as follows. Radiomics features extracted in this study are listed in Table S2. Detailed explanations of each feature can be obtained from the *pyradiomics* documentation available at http://pyradiomics.readthedocs.io/en/latest.

As for the assessment of feature extraction reproducibility, inter-class correlation coefficients (ICC) was used to assess the inter-observer reproducibility of radiomics feature extraction. We initially chose 25 random MRI images for VOI segmentation and feature extraction. The VOI segmentation was performed by two radiologists, i.e., Yong Li (reader 1) and Zhuo Wu (reader 2). The remaining image segmentation was performed by reader 1. As a result, favorable inter-observer feature extraction reproducibility was observed (ICC, median ± SD, 0.813 ± 0.172). Moreover, for the nine selected features in the radiomics signature, the ICCs ranged from 0.785 to 0.928, indicating high stability (ICC > 0.75).

(1) First-order statistics features

First-order statistics describe the distribution of the voxel intensities within the image region defined by the mask through commonly used and basic metrics. A total of 19 first-order statistics features were extracted.

(2) Shape-based features

We extracted 16 shape-based features in our study. These features include descriptors of the three-dimensional size and shape of the region of interest (ROI). They are independent from the gray level intensity distribution in the ROI and are therefore only calculated on the non-derived image and mask.

(3) Statistics-based textural features

Statistics-based textural features can reflect the homogeneity phenomenon of the images and the arrangement of the properties that change slowly or periodically on the body surface. Textural features extracted in our study included three types of matrix features: (a) 27 gray-level co-occurrence matrix (GLCM) features; (b) 16 gray-level run-length texture matrix (GLRLM) features; and (c) 16 gray-level size zone matrix (GLSZM) features. Determining the texture matrix representations requires the voxel intensity values within the volume of interest (VOI) to be discretized. Voxel intensities were therefore resampled into equally spaced bins using a bin-width of 25 Hounsfield units. This discretization step not only reduces the image noise but also normalizes the intensities across all patients, allowing for a direct comparison of all the calculated textural features between patients.

A GLCM describes the distance and angle of each pixel, which calculates the correlation between two gray levels with certain directions and distances. GLCM can reflect integrated information regarding the direction, interval, amplitude, and frequency of the images. As for GLRLM, the run length metrics quantify the gray level runs in an image. A gray level run is defined as the length in the number of pixels and of the consecutive pixels that have the same gray-level value. A GLSZM describes the amount of homogeneous connected areas within the tumor volume, of a certain size and intensity, thus reflecting the tumor heterogeneity at a regional scale.

(4) Wavelet features

Wavelet transformation effectively decouples the textural information by decomposing the original image in low- and high-frequencies. In this study, a discrete, one-level and undecimated three-dimensional wavelet transformation was applied to each MRI image, which decomposed the original image into 8 decompositions.

Consider L and H to be low-pass and high-pass functions, respectively, X to be the decomposing image, and the wavelet decompositions of X to be labeled as $X_{LLL}, X_{LLH}\, X_{LHL}, X_{LHH}, X_{HLL}, X_{HLH}, X_{HHL}, X_{HHH}$. Then, eight new images that are decomposed in three directions (x, y, z) can be obtained. Since the applied wavelet decomposition is undecimated, the size of each decomposition is equal to the original image and each decomposition is shift invariant. Thus, the original tumor delineation of the tumor volume can be applied directly to the decompositions after wavelet transformation. For each decomposition, we computed the first-order statistical features and statistics-based textural features described above, resulting in 624 wavelet features.

Reference:

1. Egger J, Kapur T, Dukatz T, et al. Square-cut: a segmentation algorithm on the basis of a rectangle shape. *PloS one* 2012; **7**(2): e31064.

II. Radiomics Score Calculation Formula

Radiomics score = 0.833385

+0.136198*Skewness

+0.555503*wavelet-HLL_Skewness

+0.000309*wavelet-LLH_Maximum

+0.000118*wavelet-LLH_glcm_ClusterShade

+0.224508*wavelet-LLH_glszm_ZoneEntropy

+12.767517*wavelet-HHH_glcm_Correlation

- 0.138853*wavelet-LLL_Kurtosis

- 9.318711*wavelet-LLL_glcm_MaximumProbability

+0.980813*wavelet-LLL_glszm_ZoneEntropy

III. Supplementary Tables

Table S1. Pathological characteristics of the patients in our study

	n (%)
Pathologic T stage, No. (%)	
pT1	10 (10%)
pT2	36 (35%)
pT3	44 (43%)
pT4	13 (12%)
Pathologic N stage, No. (%)	
pN-	74 (72%)
pN+	29 (28%)
Grade, No. (%)	
Low grade	7 (7%)
High grade	96 (93%)
Number of nodes removed	
Median	22
Range	6-59
Number of positive nodes	
Median	2
Range	1-18

Table S2. Extracted radiomics features

Groups	Subgroups	Radiomics Features
First-order statistics features		InterquartileRange, Skewness, Uniformity, MeanAbsoluteDeviation, Energy, RobustMeanAbsoluteDeviation, Median, TotalEnergy, Maximum, RootMeanSquared, 90Percentile, Minimum, Entropy, StandardDeviation, Range, Variance, 10Percentile, Kurtosis, Mean
Shape-based features		Maximum3DDiameter, Compactness2, Maximum2DDiameterSlice, Sphericity, MinorAxis, Compactness1, Elongation, SurfaceVolumeRatio, Volume, SphericalDisproportion, MajorAxis, LeastAxis, Flatness, SurfaceArea, Maximum2DDiameterColumn, Maximum2DDiameterRow
Statistics-based textural features	GLCM	SumVariance, Homogeneity1, Homogeneity2, ClusterShade, MaximumProbability, Idmn, Contrast, DifferenceEntropy, InverseVariance, Dissimilarity, SumAverage, DifferenceVariance, Idn, Idm, Correlation, Autocorrelation, SumEntropy, AverageIntensity, Energy, SumSquares, ClusterProminence, Entropy, Imc2, Imc1, DifferenceAverage, Id, ClusterTendency
	GLRLM	ShortRunLowGrayLevelEmphasis, GrayLevelVariance, LowGrayLevelRunEmphasis, GrayLevelNonUniformityNormalized, RunVariance, GrayLevelNonUniformity, LongRunEmphasis, ShortRunHighGrayLevelEmphasis, RunLengthNonUniformity, ShortRunEmphasis, LongRunHighGrayLevelEmphasis, RunPercentage, LongRunLowGrayLevelEmphasis, RunEntropy, HighGrayLevelRunEmphasis, RunLengthNonUniformityNormalized,
	GLSZM	GrayLevelVariance, SmallAreaHighGrayLevelEmphasis, GrayLevelNonUniformityNormalized, SizeZoneNonUniformityNormalized, SizeZoneNonUniformity, GrayLevelNonUniformity, LargeAreaEmphasis, ZoneVariance, ZonePercentage, LargeAreaLowGrayLevelEmphasis, LargeAreaHighGrayLevelEmphasis, HighGrayLevelZoneEmphasis, SmallAreaEmphasis, LowGrayLevelZoneEmphasis, ZoneEntropy, SmallAreaLowGrayLevelEmphasis
Wavelet features		wavelet-LLL_x, wavelet-LLH_x, wavelet-LHL_x, wavelet-LHH_x, wavelet-HLL_x, wavelet-HLH_x, wavelet-HHL_x, wavelet-HHH_x

NOTE: On the list of wavelet features, x denotes the first-order statistics features and statistics-based textural features listed above.

Table S3. Stratified analysis of the association between the radiomics signature and LN status in the combined training and validation set

		Radiomics score		
	Subgroups	**pN0**	**pN1-3**	***P***
Sex	Male	-1.402 (-1.829 to -1.007)	-0.693 (-0.971 to -0.217)	< 0.001
Age, years	<65	-1.502 (-2.018 to -1.120)	-0.693 (-0.961 to -0.160)	< 0.001
	≥65	-1.089 (-1.589 to -0.944)	-0.386 (-0.902 to -0.166)	< 0.001
MRI-reported tumor size	≤3 cm	-1.545 (-2.243 to -1.073)	-0.506 (-0.864 to -0.258)	0.005
	>3 cm	-1.259 (-1.688 to -0.969)	-0.693 (-0.961 to -0.117)	< 0.001
MRI-reported number of tumors	Single	-1.337 (-1.659 to -1.003)	-0.693 (-0.971 to -0.142)	< 0.001
	Multiple	-1.517 (-2.061 to -1.020)	-0.506 (-0.956 to -0.224)	< 0.001
MRI-reported T stage	cTa-cT2	-1.460 (-1.964 to -1.099)	-0.868 (-0.998 to -0.657)	0.003
	cT3-cT4	-1.195 (-1.684 to -0.903)	-0.469 (-0.910 to -0.142)	< 0.001
MRI-reported LN status	cN0	-1.400 (-1.832 to -1.020)	-0.694 (-0.990 to -0.177)	< 0.001
	cN1-3	-1.517 (-2.033 to -1.003)	-0.542 (-0.766 to -0.158)	0.002

NOTE: Radiomics scores are shown as medians (interquartile ranges). Stratified analysis was not conducted in the subgroup of female patients (n=14) due to the limited sample size.

IV. Supplementary Figure

Figure S1. Recruitment pathways for patients in this study

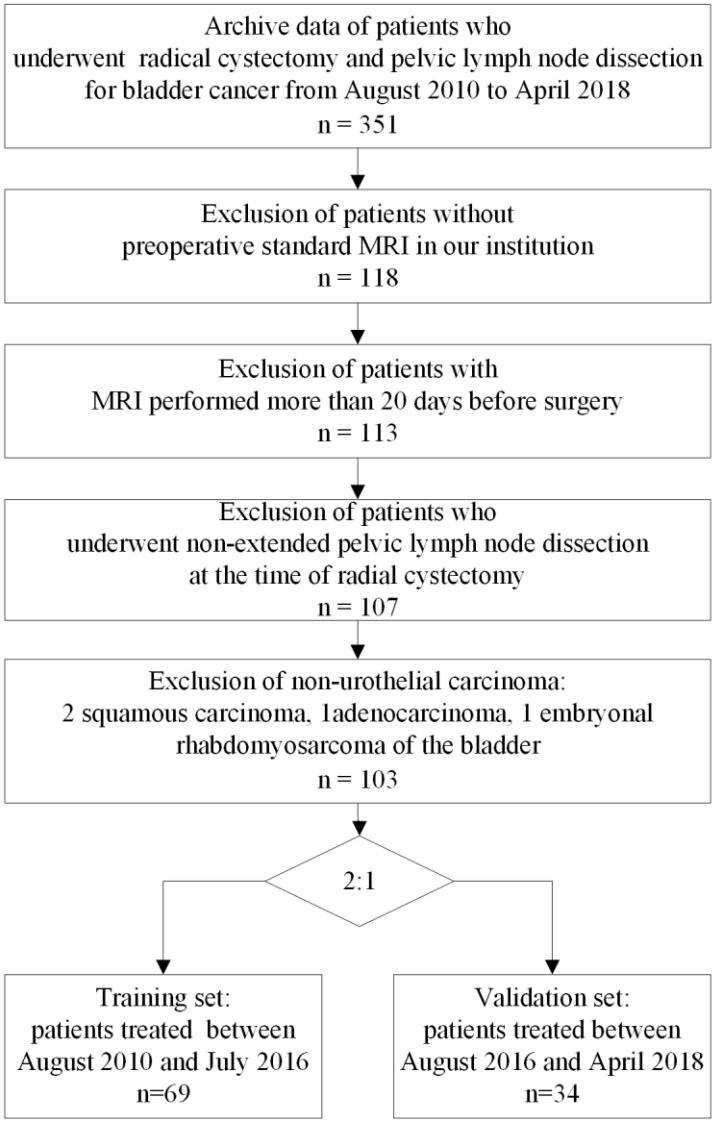

Figure S2. ROC curve analyses of the models to compare the predictive performance in the training and validation sets. (a) Training set. (b) Validation set.

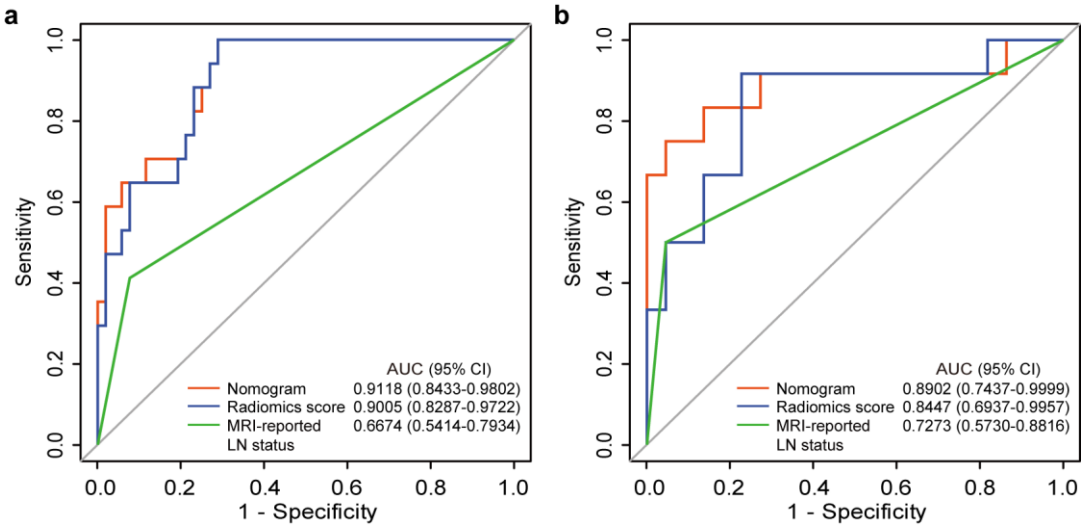

Figure S3. DCA for the nomogram in the training and validation sets. (a) Training set. (b) Validation set.